

Special Topics in Mechano-Informatics II

2018/06/27

37-176839 Koichiro Tamura

- **Difference among selective inference, multiple testing, and data splitting approaches**

- Selective inference:

A tool for correcting the selection bias when we generate the hypotheses from the data, and those hypotheses are evaluated by the same one.

- Multiple testing

Performs more than one statistical inference procedure on the same data set, causing increased significance probability if we don't apply no correction method like Bonferroni.

- Data splitting:

A sort of selective inference because of being evaluated only for the selected hypotheses from the data, but less powerful both in selection and inference.

Real world example of selection bias and discuss whether SI can solve the problem

Example: detecting trend by Social sensor from SNS, like Twitter .

The selected data could include biases because the set of people are young and familiar with tech.

SI can solve this problem if we know the distribution of the feature, like private data, social network, trust score, and so on.